

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-159865

(43)Date of publication of application : 12.06.2001

(51)Int.Cl.

G09B 5/06
 G06F 3/00
 G06F 3/16
 G06F 17/60
 G06T 13/00
 G09B 5/14
 G09B 15/00
 G09B 19/04
 G09B 19/06
 G10L 13/00
 G10L 21/06
 G10L 15/00
 G10L 15/28

(21)Application number : 2000-272404

(22)Date of filing : 08.09.2000

(71)Applicant : LUCENT TECHNOL INC

(72)Inventor : AUGUST KATHERINE G

BLACKWOOD NADINE

LI QI P

MCNERNEY MICHELLE

SHIH CHI-LIN

CHANDRASEKARAN SURENDRAN ARUN

ZHONG JIALIN

ZHOU QIRU

(30)Priority

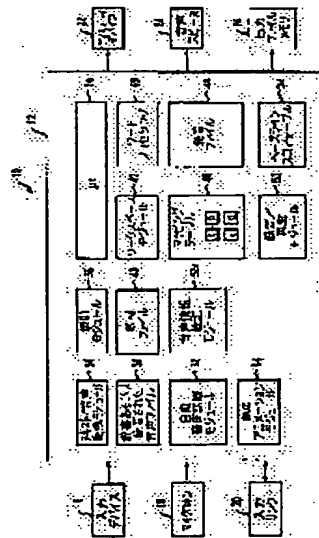
Priority number : 1999 392844 Priority date : 09.09.1999 Priority country : US

(54) METHOD AND DEVICE FOR LEADING INTERACTIVE LANGUAGE LEARNING

(57)Abstract:

PROBLEM TO BE SOLVED: To provide method and device for leading interactive language learning.

SOLUTION: In the method and device for leading the interactive language learning, a text file for processing is displayed, the basic functions for interactive learning are provided, an animation of a face is displayed, and a work space for a language construction function is provided. This system is provided with a set of language rules stored as part of a text/voice conversion sub-system, and another set of stored rules applied to a process for learning the languages. The method realized by this system includes a step for converting the text to an audible voice, a step for supplying the audible voice to a user or a student (together with an animation image when selected), a step for prompting the student so as to repeat the audible voice, a step for comparing the repeated voice of the student with the audible voice provided by the system and a step for providing a student with feedback and support, e.g. for selecting a required part and reproducing the audible voice and the voice of the student toward the student.



LEGAL STATUS

[Date of request for examination]

08.05.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C): 1998,2003 Japan Patent Office

(19)日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11)特許出願公開番号

特開2001-159865

(P2001-159865A)

(43)公開日 平成13年6月12日(2001.6.12)

(51)Int.Cl. ⁷	識別記号	F I	テリトリー(参考)
G 0 9 B 5/06		G 0 9 B 5/06	
G 0 6 F 3/00	6 5 1	G 0 6 F 3/00	6 5 1 A
	3 4 0		3 4 0 S
	1 2 8		1 2 8
G 0 6 T 13/00		G 0 6 T 13/00	A

審査請求 未請求 請求項の数19 O L (全 18 頁) 最終頁に続く

(21)出願番号 特願2000-272404(P2000-272404)

(22)出願日 平成12年9月8日(2000.9.8)

(31)優先権主張番号 09/392844

(32)優先日 平成11年9月9日(1999.9.9)

(33)優先権主張国 米国 (U S)

(71)出願人 596092698

ルーセント テクノロジーズ インコーポ
レーテッド

アメリカ合衆国, 07974-0836 ニュージ
ャーシイ, マレイ ヒル, マウンテン ア
ヴェニュー 600

(72)発明者 キャサリン ジー. オーガスト

アメリカ合衆国 07747 ニュージャーク
イ, マタワン, コロニアル ドライヴ 25

(74)代理人 100064447

弁理士 岡部 正夫 (外11名)

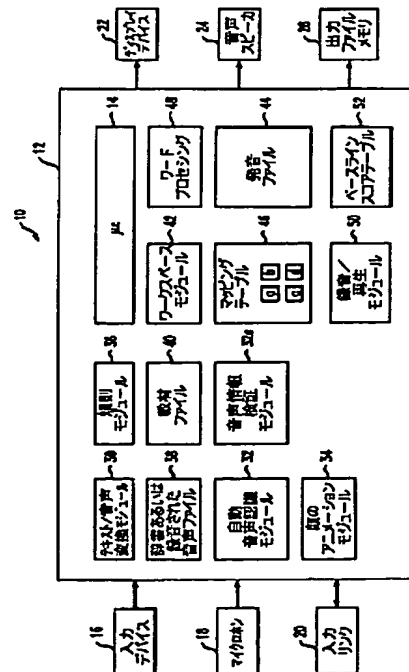
最終頁に続く

(54)【発明の名称】 対話型語学指導のための方法および装置

(57)【要約】

【課題】 本発明は対話型語学指導のための方法および装置に関する。

【解決手段】 本発明による対話型語学指導のための方法および装置においては、処理のためのテキストファイルが表示され、対話学習のための基本的な諸機能が提供され、顔のアニメーションが表示され、語学構築機能のためのワークスペースが提供される。このシステムは、テキスト/音声変換サブシステムの一部として格納されているセットの言語規則、および語学を学習する過程に適用されるもう一つのセットの格納された規則を備える。このシステムによって実現される方法は、テキストを可聴音声に変換するステップ、可聴音声をユーザあるいは学生に（選択された場合はアニメーション画像と共に）供給するステップ、学生に可聴音声を復唱するように催促するステップ、学生の復唱音声をシステムによって提供された可聴音声と比較するステップ、および、学生にフィードバックと補強を提供するステップ、例えば、必要な部分を選択し、可聴音声と学生の音声を学生に向けて再生するステップを含む。



【特許請求の範囲】

【請求項 1】 ユーザ用の対話型語学指導システムであって、
入力テキストをあるモデルに従って可聴音声に選択された言語にて変換するための第一のモジュール；前記可聴音声を復唱することを指示するプロンプトに回答してユーザによって発声された音声を受信するためのユーザインタフェース；および前記ユーザの発音を認識し、ユーザの発音と可聴音声およびモデルの一つとの比較に基づいて、ユーザにユーザが可聴音声を選択された言語にて復唱する際の正確さに関するフィードバックを供給するための第二のモジュールを備えることを特徴とする請求項 1 に記載のシステム。

【請求項 2】 さらに、前記第二のモジュールに同期された、前記可聴音声を発音する人間の顔と頭のアニメーション画像を生成するための第三のモジュールを備えることを特徴とする請求項 1 に記載のシステム。

【請求項 3】 前記人間の顔と頭のアニメーション画像が、顔と頭を透視的に描写することを特徴とする請求項 2 に記載のシステム。

【請求項 4】 前記第一と第三のモジュールが、さらに、前記ビデオ画像および可聴音声の音量、速度、および音声特性の一つを制御するためのコントロールを備えることを特徴とする請求項 2 に記載のシステム。

【請求項 5】 前記モデルが、予測モデル、音素モデル、二重母音モデル、および動的に生成されたモデルの一つから成ることを特徴とする請求項 1 に記載のシステム。

【請求項 6】 前記第一のモジュールが、前記入力テキストを構成する単語に対するモデル発音を格納するファイルを備えることを特徴とする請求項 1 に記載のシステム。

【請求項 7】 さらに、レッスンファイルを備え、前記入力テキストが前記レッスンファイル内に格納されたデータに基づいて生成されることを特徴とする請求項 1 に記載のシステム。

【請求項 8】 前記入力テキストが、システムの外部のソースから受信されたデータに基づいて生成されることを特徴とする請求項 1 に記載のシステム。

【請求項 9】 前記システムがさらに辞書ファイルを備えることを特徴とする請求項 1 に記載のシステム。

【請求項 10】 前記システムが、さらに、録音／再生モジュールを備えることを特徴とする請求項 1 に記載のシステム。

【請求項 11】 前記システムが、さらに、ワードサブグループと語彙集のワードとの間のマッピング情報を格納するテーブルを備えることを特徴とする請求項 1 に記載のシステム。

【請求項 12】 前記システムが、さらに、ワードと語彙集のワードとの間のマッピング情報を格納するテー

ルを備えることを特徴とする請求項 1 に記載のシステム。

【請求項 13】 前記システムが、さらに、ワードと音声の部分のサンプルとの間のマッピング情報を格納するテーブルを備えることを特徴とする請求項 1 に記載のシステム。

【請求項 14】 前記システムが、さらに、パンクチュエーションのテーブルを備えることを特徴とする請求項 1 に記載のシステム。

10 【請求項 15】 前記システムが、さらに、専門用語の発音ファイルを備えることを特徴とする請求項 1 に記載のシステム。

【請求項 16】 入力テキストをあるモデルを表す可聴音声に選択された言語にて変換するための第一のモジュール；前記可聴音声を発音する人間の顔と頭のアニメーション画像を生成するための前記第一のモジュールに同期された第二のモジュール；前記可聴音声を復唱することを指示するプロンプトに回答してユーザによって発声された音声を受信するように配置されたユーザインタフェース；および前記ユーザの発音を認識し、ユーザの発音と可聴音声およびモデルの一つとの比較に基づいて、ユーザにユーザが可聴音声を選択された言語にて復唱する際の正確さに関するフィードバックを供給するための第三のモジュールを備えることを特徴とするシステム。

20 【請求項 17】 音声による対話型語学指導のための方法であって、

入力テキストデータを可聴音声データに変換するステップ；前記可聴音声データに基づいて音素から成る可聴音声を生成するステップ；前記可聴音声を音声出力デバイスを通じて出力するステップ；前記可聴音声を発音する顔と頭のアニメーション画像を生成するステップ；前記可聴音声と前記ビデオ画像とを同期させるステップ；ユーザに対して前記可聴音声を復唱するように催促するステップ；前記催促に回答してユーザによって発音された音声を認識するステップ；前記可聴音声とユーザの発音を比較するステップ；および前記比較に基づいてユーザにフィードバックを供給するステップを含むことを特徴とする方法。

30 【請求項 18】 さらに、ネットワーク、格納されているレッスンファイル、スキャナ、およびインターネットの一つから前記入力テキストを受信するステップを含むことを特徴とする請求項 17 に記載の方法。

【請求項 19】 前記フィードバックが、前記可聴音声とユーザの発音の選択された部分をユーザに対して再生することから成ることを特徴とする請求項 17 に記載の方法。

【発明の詳細な説明】

【0001】

50 【発明の属する技術分野】本発明は、対話型語学指導のための方法および装置に関する。より詳細には、本発明

は、テキストファイル処理のための表示し、対話学習のための様々な機能を提供し、顔のアニメーションを表示し、語学構築機能のための学習空間を提供するマルチメディア多重モードコンピュータアプリケーションに関する。このシステムは、テキスト/音声変換サブシステムの一部として格納されているセットの言語規則、および語学を学習する過程に適用されるもう一つのセットの格納された規則を備える。このシステムによって実現される方法は、テキストを可聴音声に変換するステップ、可聴音声をユーザあるいは学生に（選択された場合はアニメーション画像と共に）供給するステップ、学生に可聴音声を復唱するように催促するステップ、学生の音声をシステムによって提供された可聴音声と比較するステップ、学生の音声に関するパフォーマンスを評価するステップ、および、学生にフィードバックおよび補強を提供するステップ、例えば、必要な部分を選択し、可聴音声と学生の音声を学生に向けて再生するステップを含む。

【0002】本発明は、とりわけ、対話型語学指導の分野に向けられ、これとの関連で具体的に説明されるが、本発明は、他の分野およびアプリケーションにおいても有益である。例えば、本発明は、一般的なスピーチ技能をスピーチの上達を目指す個人に教えるために用いることも、歌手を歌唱技能の向上を目指して訓練するために用いることもできる。

【0003】

【従来の技術】背景として、対話型語学指導プログラムが知られている。例えば、R t i s c h e v らに交付された合衆国特許第5,634,086号は、指導および評価のためのコンテキストベースの音声認識を採用する音声による語学指導方法および装置に関する。ただし、これらの既知の語学指導システムにおいては、学生がある言語を学習する際に比較のために用いるモデルとして録音された音声を用いることを強いられる。

【0004】録音された音声として教材を準備する（例えば、台本を準備する）ことに伴う作業には、語句、単語などの録音、イラスト、写真、ビデオ、その他媒体の作成、音声ファイルと画像および教材の内容との連結、コンテキストベースのレッスンのために学生との対話を複製（リプリケート）するように設計されたダイアログシステム内の代替複製（リプリカ）の多量のデータベースの用意その他が含まれる。

【0005】さらに、語学学生は、しばしば、例えば、自身の職場（例えば、コンピュータ産業、通信、自動車の修理、その他）からの技術用語など特定の分野の単語、語句、およびコンテンツの学習を希望する。このような専門分野の内容を語学教材のために録音された音声を用いて作成することは困難である。

【0006】コンテキストベースの語学学習のために録音された音声を用いる方法は、他にも多くの困難があ

る。例えば、録音媒体の品質も問題となる。これとの関連では、録音における過剰な背景雑音はその品質に影響を与える。加えて、録音された音声は、音声モデルに不当に侵入する他の多くの要因によっても影響される。例えば、録音された音声は、スピーカ（スピーカ）が特定の地理的エリアからのネイティブである場合、特定のスピーカのアクセントを含む。同様に、録音された音声は、スピーカが疲れていたり、興奮していたりなど、スピーカの特定の心理状態によって影響を受ける。このような様々な事情から、録音された音声には、学生がこれを用いて語学を学習するためには、多くの難点がある。

【0007】ユーザが、テキストのファイルを、合成あるいは録音音声技術を用いて声に出して読ませることが出来る幾つかの製品が存在する。これら製品は、通常、音声/会話変換エンジンとして知られている。これに関しては、例えば、1998年5月12日付けでM o e b i u s らに交付された合衆国特許第5,751,907号、および1998年8月4日付けでO l i v e らに交付された合衆国特許第5,790,978号を参照されたい。さらに、幾つかの現存の製品では、ユーザは、辞書に単語を追加したり、辞書内の単語の発音に修正を加えたり、テキスト/音声変換エンジンによって生成される音声を修正することができる。これに関しては、例えば、1999年4月30日付けでA u g u s t およびM c N e r n e y に代わって出願された“Graphical User Interface and Method for Modifying Pronunciation s in Text-To-Speech and Speech Recognition System s”という名称の合衆国特許出願第09/303,057号（ケース名と番号：A u g u s t 23-7）を参照されたい。

【0008】ボイスあるいはスピーチ認識システムも知られている。これらシステムは、音声パターンを認識するための様々な技法を用いる。これらには、発音の検証あるいは音声情報の検証（V I V）が含まれ、これに対しては、L u c e n t T e c h n o l o g i e s によって所有される様々な特許が適用されていたり、あるいは、これらに関してL u c e n t T e c h n o l o g i e s に様々な特許が付与されている。これら本発明と譲り受け人を同一とする特許/特許出願には、とりわけ、1996年12月20日に出願され、1998年8月18日にC h o u らに交付された合衆国特許第5,797,123号；1997年7月18日付けで出願され、B. J u a n g, C. L e e, Q. P. L i およびQ. Z h o u らに交付された合衆国特許出願896,355号（および1999年1月20日付けで公開された対応する特許出願E P 892 387 A1号）；1997年7月18日付けで出願され、B. J u a n g, C. L e e, Q. P. L i およびQ. Z h o u らに交付された合衆国特許出願897,174号（および1999年1月20日付けで公開された対応する特許出願E P 892 388 A1

号) ; および 1996 年 1 月 16 日に出願され、1997 年 7 月 15 日付けで Lee 等に交付された合衆国特許第 5,649,057 号が含まれるので、これらも参照されたい。

【0009】

【発明が解決しようとする課題】 もっぱら録音された音声に依存することなく、例えば、音声検証あるいは音声情報検証 (VIV) を組み込む高信頼音声認識技術を利用する対話型語学指導プログラムが要請されている。さらに、既知のモデルが不在の場合、予測モデルに基づいて学生の発音を評価できることが求められている。さらに、任意の音声モデルに対して信頼尺度を計算し、これからスコアを計算できるシステムが要望されている。加えて、顔のアニメーションを、学生の学習過程を支援するために、選択的に組み込むことができるシステムが要望されている。

【0010】本発明は、上述の様々な問題および他の問題を解決する新たな改善された音声による対話型語学指導システムを開示する。

【0011】

【課題を解決するための手段】 音声による対話型語学指導のための方法および装置が提供される。本発明の一面においては、システムは、入力テキストを可聴音声に選択された言語にてデジタル的に変換するための第一のモジュール、可聴音声の復唱を試みるユーザの音声を受信するように配置されたユーザインタフェース、ユーザの発音を認識し、ユーザに、ユーザの発音と前記可聴音声、任意の音声モデル、予測モデル、音素モデル、二重母音モデル、あるいは動的に生成されたモデルとの比較に基づいて、ユーザが前記可聴音声を選択された言語にてどの程度正確に真似ることができるかのフィードバックを供給するための第二のモジュールを備える。

【0012】本発明のより限定された一面によると、さらに、前記第二のモジュールに同期された、可聴音声を発音する人間の顔と頭のアニメーション画像を生成するための第三のモジュールが設けられる。

【0013】本発明のもう一面によると、前記人間の顔と頭のアニメーション画像は、顔と頭を透過的に描写し、画像は、あらゆる角度から見られるように、回転、傾斜、その他の操作を加えることができる。

【0014】本発明のもう一面によると、前記第一と第三のモジュールは、さらに、ビデオ画像および可聴音声の音量、速度、および音声特性の一つを制御するためのコントロールを備える。

【0015】本発明のもう一面によると、前記モデルは、予測モデル、音素モデル、二重母音モデル、および動的に生成されたモデルの一つから成る。

【0016】本発明のもう一面によると、前記第一のモジュールは、入力テキストを構成するワードあるいはサブワードに対するモデル発音を格納するファイルを備え

る。

【0017】本発明のもう一面によると、システムは、レッスンファイルを備え、前記入力テキストはこれに基づいて生成される。

【0018】本発明のもう一面によると、前記入力テキストは、システムの外部のソースから受信されたデータに基づいて生成される。

【0019】本発明のもう一面によると、システムはさらに辞書ファイルを備える。本発明のもう一面によると、システムはさらに録音／再生モジュールを備える。

【0020】本発明のもう一面によると、システムはさらにワードのサブグループと語彙集のワードとの間のマッピング情報を格納するテーブルを備える。

【0021】本発明のもう一面によると、システムはさらにワードと語彙集のワードとの間のマッピング情報を格納するテーブルを備える。

【0022】本発明のもう一面によると、システムはさらにワードと音声の部分のサンプルとの間のマッピング情報を格納するテーブルを備える。

【0023】本発明のもう一面によると、システムはさらにバンクチューエーションのテーブルを備える。

【0024】本発明のもう一面によると、システムは、サブワードと別の言語での対応するサブワードのテーブルを備える。例えば、ワードの音の訓練では、(第二の言語を母国語とする学生が) 第一の言語を学習する場合、第一の言語からのサブワードが、学生に音の類似する比較を示すために、第二の言語におけるサブワードにマッピングされる。このサブワードのテーブルを用いると、さらに、いずれの言語からも、サブワードを用いて、語彙集のワードを検索し、表示／再生することが可能となる。

【0025】本発明のもう一面によると、音声による対話型語学指導のための方法が提供される。この方法は、入力テキストデータを可聴音声データに変換するステップ、可聴音声データに基づいて音素あるいは二重母音から構成される可聴音声を生成するステップ、可聴音声を発音する顔と頭のアニメーション画像を生成するステップ、可聴音声とビデオ画像とを同期させるステップ、ユーザに対して可聴音声を復唱するように催促するステップ、催促に応じてユーザによって発音された音声を認識するステップ、前記可聴音声の音素あるいは二重母音とユーザの発音とを比較するステップ、および比較に基づいてユーザにフィードバックを供給するステップを含む。

【0026】本発明のもう一面によると、ある言語の基本的な音素と二重母音のインベントリ (収集) を表す一連の文が提供される。学生は、これら文を朗読することを求められ、学生の音声は録音される。サブワードが解析され、学生のベースラインスコアと、開始時のパフォーマンスが決定される。これが、進歩を測定するため、

訓練のためのレベルを設定するため、あるいは学習すべき領域を決定するために用いられる。

【0027】本発明のもう一面によると、ある与えられた学生の集団から構成される語学クラスの評価レベルに対する基準スコアのテーブルが提供される。学生の進歩の測定および評価は、個人ベースにて行なうことも、選択された集団との比較で行なうこともできる。

【0028】本発明のもう一面によると、学生の発音に対するスコアが、サブワード、ワード、文、あるいはパラグラフの単位で提供される。学生は、総合スコアを得ることも、音声の個々の部分に関するスコアを得ることもできる。

【0029】本発明のもう一面によると、音声の検証に関する正規化問題は、インタフェースを介して管理される。異なる持続時間および複雑さの音声を、画面上のアニメ化されたカーソルを用いて、システムによって、あるいは学生によって、設定することができ、学生がアニメ化されたカーソルに沿って読み進むと、検証過程によってハイライトされたスキストが解析されるべき音声ファイルと対応付けられる（学生の発音と音声ファイルとが比較される）。

【0030】本発明のもう一面によると、ある与えられた言語の既知のサブワードあるいはワードに対して、ある種の録音された音が、自然な音を強調するためにさしはさまれる。これらワードは、事前に録音された辞書やアプリケーションから取ることも、他のソースから取ることもできる。

【0031】本発明のもう一面によると、ベースラインスコアがテーブル内に記録される。このテーブルは、学生に対して選択されるべき適当なレベルのレッスンを決定するために用いられる。このシステムを用いることで、システムは、同一のテキスト、内容等を、信頼尺度の閾値を変更することで、異なる能力を持つ学生に対して自動的に用いることが可能となる。

【0032】本発明のもう一面によると、先生あるいは学生は、グラフィカルユーザインタフェースを用いて、信頼尺度、等級レベル、その他の属性に対する閾値を設定あるいは修正することができる。

【0033】本発明のもう一面によると、学生は、識別を登録、ベースラインスコア、およびその後のレッスンのスコアを入力することで、カスタム化されたレッスンを実現したり、進歩を追跡することができる。

【0034】本発明の適用のさらなる範囲が以下の詳細な説明から明らかになるものである。ただし、詳細な説明および特定な実施例は、本発明の好ましい実施例を示すため、および単に解説のために与えられているものであり、当業者においては本発明の精神および範囲から逸脱することなく、様々な変更および修正が可能であることに注意する。本発明は、デバイスの様々な要素の構造、配列、および組合せ、並びに、方法のステップから

構成され、本発明の目的は、特許請求の範囲、以下の詳細な説明、および付録の図面において、より詳細に定義され、具体的に説明され、図解されるやり方によって達成されるものである。

【0035】

【発明の実施の形態】次に、図面の説明に移るが、これら図面は単に本発明の一つの好ましい実施例を示すもので、本発明を制限するものではないことに注意する。図1は、本発明による好ましいシステムの概要を示す。図1に示すような対話型語学指導システム10が開示される。システム10は、マイクロコントローラあるいはマイクロプロセッサ14を備えるコンピュータ化された装置あるいはシステム12を備える。システム10は、さらに、キーボード、マウス、その他の一つあるいは複数から成る入力デバイス16、マイクロホン18、入力リンク20、一つあるいは複数のディスプレイデバイス22、音声スピーカ24および出力ファイルインタフェースユニット26を備える。これら全ての要素は従来のもので、当業者においては周知であり、ここでの説明は割愛する。さらに、システム10は、適当な形態にて、クライアントサーバおよびスリムクライアントアーキテクチャ内に組み込むことも、あるいはこれと互換性をもたすこともできる。さらに、このシステムはコンピュータディスクやインターネットを通じて提供あるいは供給し、さらに、より小さなデバイスあるいは携帯デバイスにダウンロードすることもできることに注意する。

【0036】システム12は、様々な要素を備える。これら要素は、図1に示すように、コンピュータシステム12内に組み込むことも、あるいは、コンピュータシステム12の遠隔に配置し、ネットワークあるいは他の接続を通じてアクセスすることもできる。図1に示すように、システム12は、テキスト/音声変換（TTS）モジュール30と、自動音声認識（ASR）モジュール32を備える。これらモジュールは、従来のもので、当業者においては周知である。好ましくは、テキスト/音声変換（TTS）モジュール30は、例えば、1998年5月12日付けでMoebiusらに交付された合衆国特許第5,751,907号および1998年8月11日付けでOliveらに交付された合衆国特許第5,790,978号の教示を組み込み、自動音声認識（ASR）モジュール32（音声情報検証部分32aも含めて）は、例えば、1996年12月20日に出願され、1998年8月18日付けでChouらに交付された合衆国特許第5,797,123号；1997年7月18日付けでB. Juang, C. Lee, Q. P. Li, およびQ. Zhouによって出願された合衆国特許出願第896,355号（および1999年1月20日付けで公開された対応する出願EP 892 387 A1）；1997年7月18日付けでB. Juang, C. Lee, Q. P. Li, およびQ. Zhouによつ

て出願された合衆国特許出願第897, 174号(および1999年1月20日付けで公開された対応する出願EP 892 388 A1); および1997年7月15日付けでLeeらに交付された合衆国特許第5, 649, 057号の教示を組み込む。詳しくは、これら特許を参照されたい。テキスト/音声変換(TTS)モジュール30は、デジタルデータとして格納されているテキストをスピーカ24によって音素の形式にて出力するために音声信号に変換し、自動音声認識(ASR)モジュール32は、マイクロホン18を通じて受信される音声信号をデジタルデータに変換する。

【0037】テキスト/音声変換(TTS)モジュール30は、これに関連して、テキストから可聴音声への変換に用いる規則モジュール36を備える。より具体的には、規則モジュール36は、その内部に可聴音声への変換が求められている単語を多重レベル解析するためのコードを格納する。これら規則モジュールは、次々と、選択された単語を解析し、単語を文のレベルで解析し(例えば、付近の単語や音声の一部を解析し(例えば、“address”が名詞であるか動詞であるか決定し))、次に、文の形式を解析する(例えば、その文が疑問文であるか、平叙文であるか決定する)。この解析スキームは、各単語をその単語が用いられている文脈内でより正確に発音する(例えば、正しく強調する)ことを助ける。テキスト/音声変換(TTS)モジュール30は、さらに、選択された単語の正しい発音を助けるための辞書あるいは録音された音声ファイル38と接続されていると共に、勿論、教材ファイル40にも接続されており、これからレッスン用のテキストを検索できるようになっている。教材テキストは、インターネット、LAN、WAN、スキャナ、クローズドキャプションデバイス(閉鎖字幕デバイス)などを含む様々な他のソースから入力リンク20を通じて得ることもできる。この機能は教材内容をシステムの機能から分離させることを可能にする。つまり、本発明によるシステムおよび方法は、ユーザあるいは学生の必要性および/あるいは要求に合わせて様々な異なる教材内容に適用することができる。

【0038】一つの好ましいテキスト/音声変換(TTS)モジュールあるいはエンジンは、その内部に、テキストとして入力された全ての単語あるいはサブワードのモデム発音を備える。これらモデルファイルは、最終的には、後に詳細に説明するように、学生の発音と比較するために用いられる。本発明によるシステムでは、辞書あるいはファイル内の任意の単語を用いることができる。他の語学学習製品では、教材は録音された単語あるいは語句に制限されているが、本発明の好ましいテキスト/音声変換(TTS)モジュールは、テキストあるいはテキストファイルを認識・処理し、これを可聴音声に変換する能力を備え、テキスト/音声変換(TTS)モ

ジュールを追加することで、教材内容をより柔軟なものにすることができる。つまり、テキスト/音声変換(TTS)モジュールを用いることで、本来は他の目的を持つ材料、現在起こっている事件、ニュース、ウェブ内容、ファイル、専門記事などを、学習のための特別の目的に合わせた教材、例えば、音声セラピー用の教材に加工することが可能となる。

【0039】好ましいテキスト/音声変換(TTS)モジュールは、語学学生に対して、テキストをその言語の音声規則に従って模範として発音してみせる。テキスト/音声変換(TTS)モジュールの音声品質は極めて高い。このため、学生は、ネイティブな英語を話す先生を必要とせず、また、レッスンに用いる単語を事前に録音し、システム内に格納しておく必要もなしに、辞書あるいはファイル内の任意の単語の発音を聴くことができる。このシステムは単語の抑揚および調子を文脈に依存して変化させることができるが、録音された音声ではこれを扱うことは困難である。さらに、テキスト/音声変換(TTS)モジュールに特別な発音ファイルを追加することで、地域によるアクセントの違いその他を試みることもできる。例えば、ある単語をアメリカの標準音で発音させたり、ある地域の方言で発音させたり、スペイン語を英語で発音させたり、正式な名前、商標、専門用語の発音を確かめたりすることもできる。

【0040】自動音声認識(ASR)モジュール32は、自動音声認識(ASR)モジュール32に発音の検証を提供する音声情報検証(VIV)部分32aを備える。音声情報検証(VIV)部分32aを備える自動音声認識(ASR)モジュール32のこの好ましい実現は、テキスト/音声変換(TTS)モジュールによって処理された音素の出力、音声、自身の音声モデル、あるいは任意の生成された音声モデルと学生の発音を比較する。音声情報検証(VIV)部分32aは、学生の発音とテキスト/音声変換(TTS)モジュールによって生成されたファイルがどの程度一致するか解析する。この比較は、学生へのフィードバックの基礎となる。総合スコアがフィードバックとして学生に供給される。加えて、個々の単語の部分あるいは音素を個々に比較することで、学生が具体的にどの部分に発音の問題を抱えるかが示される。フィードバックが生成され音声の各部分に対して学生に供給される。言語の規則、識別されたワードあるいはワードセグメントの識別、学生の母国語からもたらせられる既知の発音問題、学生の達成レベルなどに基づいて学生に発音の補強が提供される。

【0041】アニメーションモジュール34は学生を視覚的に支援する。モジュール34は、テキスト/音声変換(TTS)モジュール30と同期されており、テキストファイルを検索し、テキスト/音声変換(TTS)モジュールあるいはエンジンと協力して、人の頭と顔のアニメーション画像を通じて、学生に単語を発音する様子

を示す。好ましくは、人の頭と顔のアニメーション画像は、三次元にて表現され、この画像は、様々な角度から見られるように、回転したり、傾けたりできるようにされる。こうして、学生は、発音の際の顔や口の動きや、舌や唇や歯の位置を見ることができる。アニメーションモジュールは、顔の動きを、当業者においては周知のやり方で、テキスト／音声変換（TTS）モジュール30による処理と同期させる。学生は、アニメーション画像、すなわち先生を任意の角度から見ることができ、必要であれば、透過モードにて歯と舌の位置を詳しく調べることもできる。先生の画像は修正することもできる。例えば、先生の音量、速度、音声特性は、学生によってコンピュータインタフェースを通じて変更することができる。例えば、音声は、男性の声にしたり、女性の声にしたり、高くしたり、低くしたり、早くしたり、遅くしたり、変えることができる。後に詳細に説明するように、言語の規則、学生の母国語からもたらせられる既知の発音問題、学生の熟練レベルなどに基づいて学生に補強を提供することもできる。

【0042】システム10は、さらに、ワークスペースモジュール42を備えるが、これは、ディスプレイデバイス22上に視覚ワークスペースを生成するために用いられる。ワークスペースモジュール42は、発音モジュール44、マッピングテーブル46、ワードプロセッシングモジュール48、および録音／再生モジュール50と連動されている。

【0043】発音モジュール44は、ワード、ワードサブグループ、およびある言語の典型的な音を教えるために用いられる語彙集ワードのレコードを含むデータベース、ある単語がある文脈下でどのように発音されるかを教えるために用いられる音声の部分の模範、およびパンクチュエーションのテーブルを備える。サンプル単語が選択され、言語の文法および言語学上の規則に基づいて発音データベースが形成される。好ましくは、各文字あるいは文字グループ（二重母音）に対するサンプル単語が、その文字の発音における典型的な事例からあまり一般的でない事例の順に並べられる。発音モジュール44は、さらに、特別な発音ファイルを追加することで、アクセントの地方による違いを示すこともできる。これは、特定の文脈での特徴を示すために用いることができる。例えば、これを用いて、これに限定されるものではないが、ある単語をアメリカの標準音で発音させたり、ある地域の方言で発音させたり、スペイン語を英語で発音させたり、正式な名前、商標、専門用語の発音を確かめたりすることもできる。

【0044】マッピングテーブル46は、ワードサブグループとある言語の典型的な音を示すために用いられる語彙ワードとの間のマッピングを格納するテーブル46a、ワードとある言語の典型的な音を示すために用いられる語彙ワードとの間のマッピングを格納するテーブル

46b、およびワードとある単語がある文脈下でどのように発音されるかを示すために用いられる音声の部分からのサンプルとの間のマッピングを格納するテーブル46cを備える。このシステムは、さらに、ある言語において典型的に用いられるパンクチュエーションのサンプルを格納するテーブル46dを備えるが、これは、レッスンにおいて独立的に用いることも、サブワード、ワード、あるいはグループの背景内で用いることもできる。

【0045】図2は、システム10によって生成される主ウィンドウがユーザによってどのように見えるかを概略的に示す。ウィンドウ60は、ワークスペースモジュール42と関連するワークスペース62、主として自動音声認識（ASR）モジュールと関連するフィードバックエリア64、主としてアニメーションモジュールと関連するアニメーションエリア66および主としてテキスト／音声変換（TTS）モジュールおよびアニメーションモジュールと関連する制御エリア68を備える。ワークスペース62は、学生に対して教材テキストを表示するため、およびこれを操作するために用いられる。フィードバックエリア64は、システムによって学生に提供されるフィードバックを表示および操作するために用いられる。これについては、後により詳細に説明する。アニメーションエリア66は、図示するように、アニメ化された顔と頭66aのサンプルを示す。最後に、制御エリア68は、音量調節68a、速度調節68b、停止ボタン68c、再生ボタン68d、休止ボタン68e、および録音ボタン68fなどのユーザインタフェース制御アイコンを備える。学生は、対話的にウィンドウ60を操作することで、本発明による機能を遂行することができる。

【0046】図3は、本発明の方法の好ましい実施例の概要を示す。図3並びに図4～11との関連で説明する方法は、以下の説明を一読することで当業者においては容易に明らかとなるハードウェアおよびソフトウェア技法を用いて実現されることに注意する。

【0047】図3は、本発明の方法の概要を示す。方法300は、ステップ302から開始され、ここで、テキストが入力され、次に、ステップ304において、入力されたテキストが可聴音声データに変換される。次に、ステップ306において、可聴音声データに基づいて可聴音声が生産され、出力される。勿論、可聴音声は、予測モデル、音素モデル、二重母音モデル、動的に生成されたモデルなどを含む様々なモデルに基づいて表すことができる。これらモデルは、主として自動音声認識（ASR）モジュールと関連する要素によって生成される。ただし、幾つかの状況においては、テキスト／音声変換（TTS）モジュールを用いて、音声モデルが生成される。次に、学生によって要求される場合は、ステップ308において、人の顔と頭のアニメーション画像が主として顔面アニメーションモジュール34によって生成さ

れ、ステップ310において、可聴音声とアニメーション画像が同期される。次に、ステップ312において、学生が、可聴音声を復唱するように、つまり、単語を真似て発音するように催促される。次に、ステップ314において、システムが学生の発音を認識し、ステップ316において、学生の発音が、主としてモジュール32(音声情報検証部分32aも含む)を用いて可聴音声データと比較される。次に、ステップ318において、比較と信頼尺度に基づいてフィードバックが学生に供給される。信頼尺度はカスタム化されたスコアテーブルと関連付けられ、当分野においては周知のように様々なやり方で、キャリブレーションポイントとして用いられる。フィードバックは、好ましくは、ユーザが可聴音声を選択された言語にてどれだけ上手に復唱できるかの精度を反映する。

【0048】図4は、本発明の方法のテキストの選択および可聴音声生成過程、すなわち、ステップ302、304、および306をより詳細に示す。方法400のステップ402において、入力テキストが選択され、ステップ404において、ユニバーサル資源ロケータ(URL)を用いるか、格納されたフィールドを用いるかが決定される。URLを用いる場合は、ステップ406において、URLをフィールドにタイプ入力することで、テキストが検索される。ファイル内のテキストを用いる場合は、ステップ408において、ファイルが選択される。いずれの場合も、ステップ410において、検索されたテキストがワークスペース62内に表示される。次に、ステップ412において、再生ボタン68dが押されるか、あるいは“クリック”される。次に、ステップ414において、選択されたテキストのソースがURLを用いて探索されたものであるか、ファイルからのものであるか決定される。テキストがURLを用いて得られた場合は、ステップ416において、マークアップ言語が予備処理される。テキストは、テキスト/音声変換(TTS)モジュールによる処理のための理想的な形式を得るために予備処理される。例えば、マークアップ言語やイラストの除去や、電子メールやファクシミリなどの既知のあるいはプロビジョンドフォーマットの構文解析が行なわれる。いずれの場合も、次に、ステップ418において、テキストのサブセットが先取りされ、ステップ420において、テキストから音声への変換が開始される。オプションとして、ステップ422において、音声の速度がチェックされ、ステップ424において、音量がチェックされる。次に、ステップ426において、音声再生が完了したか否か決定される。可聴音声の再生が完了していない場合は、ステップ418~428が反復される。可聴音声の再生が完了した場合は、ステップ430において、この過程は終了する。

【0049】図5は、アニメーション(例えば、先生の

プロンプトの表示)が用いられる場合の図3のステップ302~310を詳細に示す。簡素化の目的で、方法500は、単に、図4のステップ418~430の部分に追加として示される。

【0050】図4のステップ412において再生ボタン68dが押された後、図5のステップ502において、選択されたテキストのサブセットが先取りされる。次に、ステップ504において、テキストから音声への変換が開始される。ステップ506において、テキストからアニメーションへの処理も開始される。次に、必要であれば、ステップ508において速度がチェックされ、ステップ510において、音量がチェックされる。次に、ユーザに、ステップ512において、音声再生が再生され、ステップ514において、顔の動きがアニメーションエリア66に出力される。次に、ステップ516において、アニメーションを伴う可聴音声の再生が完了したか否か決定される。完了していない場合は、ステップ502から516が反復される。アニメーションを伴う可聴音声の再生が完了した場合は、ステップ518において、この過程は終了する。

【0051】図3に戻り、ステップ312において、学生が再生された可聴音声を復唱するように催促される。これとの関連で、システムは、学生に、先生の模範を復唱するように、積極的に催促することも、消極的にすることもできる。このプロンプトの方法には、カーソルを移動する方法、テキストのハイライトされたエリアを移動する方法、アイコンを移動する方法などが含まれる。可聴プロンプトには、これに限定されるものではないが、“私についてきて”と述べた後に、復唱すべき単語を表示するプロンプトが含まれる。このプロンプトの速度は、調節可能である。

【0052】学生は、レッスンの際に、自身の発音を録音することを選択し、先生の発音と自身の発音をサイド・バイ・サイドに聴き、比較することもできる。後に説明するように、録音された発音を自動音声識別機能に掛けることで、学生のパフォーマンスを評価することもできる。

【0053】図6は、音声を録音する過程を詳細に示す。この過程は、ステップ602において、再生ボタン68bか、録音ボタン68fを選択すること、つまり、これを押すあるいは“クリック”することで開始される。次に、ステップ604において、求められているテキストファイルが格納されているものであるか、あるいはURLを用いて検索されるべきものであるか決定されている。URLが用いられた場合は、ステップ606において、マークアップ言語が予備処理される。次に、ステップ608において、テキストが先取りされ、ステップ610において、テキストがハイライトされる場合は、その部分が選択される。いずれの場合も、次に、ステップ612において、テキストから音声への変換が開始さ

れ、ステップ614において、アニメーションを使用するか否か決定される。アニメーションを必要とする場合は、ステップ616において、アニメーションデータが処理される。アニメーションが処理されるか否かに関係なく、ステップ618において速度がチェックされ、ステップ620において音量がチェックされる。次に、ステップ622において音声再生され、必要であれば、ステップ624においてアニメーションが表示される。次に、ステップ626において、再生が完了したか否か決定される。処理が完了していない場合は、ステップ606～626が反復される。

【0054】再生が完了した場合は、図7のステップ702において、学生が催促される。ステップ704において、学生が復唱すべきテキストがハイライトされる。ステップ706において、速度がチェックされる。ステップ708において、学生の録音が開始され、ステップ710において、カーソルが指定された速度にて移動される。次に、ステップ712において、過程が完了したか否か決定される。完了していない場合は、ステップ702～712が反復される。録音の過程が完了した場合は、この過程はステップ714において終了する。

【0055】再び図3に戻り、システムは、ステップ314において、音声を認識し、ステップ316において、学生の発音が、可聴音声ファイルあるいはレコードからのモデルと比較され、ステップ318において、フィードバックが供給される。ステップ316との関連で、当業者においては容易に理解できるように、学生の発音は、可聴音声、任意の音声モデル、予測モデル、音素モデル、二重母音モデル、動的に生成されたモデルのいずれと比較することもできる。フィードバックは、様々なやり方で提供することができ、一つのフィードバックの形態においては、学生に向けて自身の録音されたレッスンを再生される。

【0056】図8は再生過程をより詳細に示す。最初に、この方法800のステップ802において、先生を再生するか、学生を再生するかが決定される。先生が選択された場合は、ステップ804において、テキストがハイライトされ、ステップ806において、速度がチェックされ、ステップ808において、音量がチェックされ、ステップ810において、テキストから音声への変換が処理され、ステップ812において、アニメーションが処理され、ステップ814において、音声再生され、ステップ816において、アニメーションが動かされる。次に、ステップ818において、再生が完了したか否か決定される。完了していない場合は、ステップ804～818が反復される。過程が完了している場合は、ステップ820においてこの過程は終了する。

【0057】ステップ802において学生が選択された場合は、ステップ822において、再生されるべきテキストがハイライトされ、ステップ824において速度が

チェックされ、ステップ826において音量がチェックされ、ステップ828において、学生の音声、すなわち、録音された音声再生される。次に、ステップ830において、再生が完了したか否か決定される。完了していない場合は、ステップ822～830が繰り返される。学生の再生が完了した場合は、ステップ832においてこの過程は終了する。

【0058】学生は、自身の発音がどの程度先生のモデル発音と一致するかを知るために評価することを選択することもできる。ワード、ワードセグメント、文、あるいは文のグループがどの程度正確に発音されたかを調べるためには、自動音声認識（発声の検証）および音声情報検証（VIV）モジュールが、モジュール30、32、32a（および関連する要素）と組み合わせて用いられる。一つの好ましい形態においては、発音の検証においては、“P”と“B”あるいは“T”と“D”などの破裂音が識別される。正確さのスコアリングには、これら限定されるものではないが、全体としてのパフォーマンスに関するグロススコア、個々の文に関するスコア、個々の単語に関するスコア、および音素に関するスコアが含まれる。

【0059】学生へのフィードバックは様々な形式を取ることができる。これらフィードバックは、パフォーマンスを測定するために用いることも、補強を決めるために用いることも、このアプリケーションの機能レベル（初級者、中級、上級）を決めるために用いることもできる。フィードバックは、明示的に与えることも、あるいは、ファイルとして出力ファイルメモリ26（図1）を通じて先生に送ることも、あるいはこの両方とすることもできる。

【0060】総合スコアには、（文のグループ、文、ワード、サブワードに対する）数値が含まれ、これら数値は、初級、中級、上級、学習歴などの学生のレベルや、参照ファイルとして用いられる音節、その他に合わせて変えることができる。システムは、このスコア情報をフィードバックエリア64内に表示するように設定することも、あるいはこのスコア情報は学生には表示しないこともできる。アプリケーションは、学生のレベルを、システム内に含まれる統計情報、ネットワークを通じて得られる情報、学生がシステムと対話している際に収集された学生固有の情報、あるいは、学生の自己査定や先生もしくはセラピストによる査定に基づいて決定する。学生へのフィードバックに、パフォーマンスのレベルを示すアイコンを用いることもできる。例えば、これに限定されるものではないが、グループに配列された星、円その他の一連のシンボルを用いてパフォーマンスを示すことも、あるいは、タスクが上手に達成されたことを示すためにしばしば用いられる様々なシンボルの任意の一つを用いることもできる。一例として、3つの円を用意し、学生がもう少し練習を要する場合は、2つ示し、モ

デルと一致するためにはかなりの練習を要する場合は、1つ示し、(学生のレベルから見て)モデルとかなり一致し、それ以上の訓練は必要としない場合は、3つとも全部示すこともできる。パフォーマンスのレベルを示すためにカラーを用いることもできる。

【0061】パフォーマンスに関するフィードバックは、学生がテキストを読んでいる際に与えることも、タスクを終了した時点で与えることもできる。学生がテキストを読んでいる最中に音声情報検証処理(あるいは発音検証処理)を用いて、パフォーマンスに関するフィードバックをリアルタイムにて表示することもできる。システムは、パフォーマンスを示すために、例えば、これに限定されるものではないが、バー、アイコン、カラー、音響効果、TTSテキストファイルなどを含むグラフィックや音響キューなどを任意の方法を用いることができる。システムは、学生に問題が存在する領域を指摘し、学生は、必要であれば、練習を反復したり、速度を変更したり、別のモデルや機能、例えば、単語構築機能に移ったり、先生の模範を再び聴いたりすることができる。特定のパフォーマンス問題に基づいてデフォルトオプションが確立され、例えば、VIV機能によって決定される。

【0062】いったん学生がある程度の期間練習すると、学生は、再びフィードバックをリクエストすることができる。フィードバックとして、学生に、これに限定されるものではないが、例えば、テキスト内のハイライトされた単語、総合スコア、練習の個々のセグメントに対する個々のスコア、全体としての達成度を示すアイコン、パフォーマンスを示す可聴フィードバックなどを含む要約が供給される。可聴フィードバックは、かつさいする群衆などの音響効果とすることも、上手に発音できなかった単語にカーソルが移動されたとき、かん音を響らせることもできる。学生は、再生機能を用いて、モデルと自身の発音を比較することもできる。

【0063】発音スキルを訓練したり、定義を再検討したりするために、単語およびパンクチュエーションのリストを用いることができる。このリストは、教材ファイル(例えば、教材ファイル40)から作成することも、辞書や他の参考資料から作成することも、インターネットから(例えば、入力リンク20を介して)作成することも、学生によって誤って発音された(例えば、発音ファイル44内に格納されている)サブワード、ワードあるいはグループのリストから作成することもできる。このシステムは、一つの長所は、ワードとパンクチュエーションによって影響される発音とを組み合わせることができ、システムは、これら組合せに対してフィードバックおよび補強を提供することができることである。誤って発音された単語が含まれる場合、システムは、これら単語を上手に発音できたものからあまり上手に発音できなかったものへと順番に並べることも、デフォルトとし

て、最も訓練を必要とする項目をリストの先頭に与えることもできる。単語リストはワーキングウインドウ内に表示される。説明の例では、ワーキングウインドウがフィードバックエリア64内に表示される。学生は、ツールを用いてこのリストをスクロールさせることができる。学生は、音声コマンドあるいはマウスを用いて、ある単語をハイライトさせ、選択することができる。音楽あるいはワード(あるいはグループ)がハイライトさせることで、その単語を発音する先生の音声を聴くこともできる。オプション機能として、サブワード、ワードあるいはグループをハイライトさせたとき、システムが、先生の発音と学生の発音を、サイド・バイ・サイドにフィードバックするようにすることも、録音された辞書に行き、音声ファイルを再生するようにすることもできる。学生は、この時点で、再びその単語を発音し、フィードバックを得ることもできる。単語を選択すると、学生に対して、ウインドウ内に、より詳細なワークスペース機能が表示される。ワークスペース機能は、言語規則を用いて、サブワード、ワード、あるいはグループを処理し、スペル、パンクチュエーション、アクセント、音節、その他を表示する。学生は、ある模範サンプルを選択して、それを再び聴き、その発音を再び試みることもできる。学生は再び評価され、パフォーマンスが改善され学生あるいは先生によって満足できるフィードバックが得られた場合は、その単語の訓練は終了する。満足できない場合は、システムによってさらなる支援が与えられる。

【0064】学生が可聴音声およびフィードバックされた単語を発音するのに問題を持つ場合は、ワーキングウインドウ60を通じて補強が提供される。表示されたサブワード、ワード、グループの部分にカーソルを移動させると、それを発音するデフォルト機能が起動される。この機能はターンオフすることもできる。ワードの部分を選択すると、学生を助けるための補強ウインドウが提供される。この補強ヘルプの一例としては、“Try this... in the word ‘graduate’ the ‘d’ is pronounced with a ‘j’ sound as in ‘jar’ (これを試みて下さい。単語‘graduate’における‘d’は、‘jar’における‘j’音と同様に発音される)”といったメッセージが含まれる。発音のための既知の言語規則を示すテーブルの助けをかりて、選択されたサブワード、ワード、あるいはグループにおける事情に基づいてテキストメッセージが自動的に作成される。学生は、このメッセージをウインドウ内に見ることも、先生がそのメッセージを発音するのを聴くこともできる。

【0065】メッセージはシステムによってネスト化され、特定のサブワード、ワードあるいはグループケースに対して、モデルと学生との間の一致を困難にさせる複

数の言語学的理由が存在する場合は、メッセージが学生に事前に決められた順番に提示される。この場合、最初に、最も一般的な発音規則が提示され、次第に、より稀な規則が提示される。

【0066】発音上のエラーに関する既に知られている知識が学生を支援するために用いられる。例えば、英語を学習する韓国人が不得意とする典型的なエラーが知られている。“Try this...”なるメッセージ系は、システムのユーザを考慮し、特定の学生に、学生の自己同一（国籍などに）基づいて最も適した指導を与える。この機能に対するテキストあるいは可聴ヘルプは、母国語で提示することも、目標言語にて提示することも、これらを組合せて提示することも、あるいはこれらの両方にて提示することもできる。例えば、発音ファイル44は、サブワードと別の言語での対応するサブワードのテーブルを含めることもできる。例えば、ワードサウンドドリル（単語の音の訓練）の場合、（第二の言語を母国語として話す学生が）第一の言語を学習する場合、第一の言語からのサブワードが、学生にサウンドの類似する比較を与えるために、第二の言語のサブワードにマッピングされる。さらに、サブワードテーブルを用いて、そのサブワードを含む語彙集ワードを識別し、これをどちらかの言語にて表示／再生することもできる。ワークスペース（学習空間）と関連するもう一つの練習機能として、練習するサブワード、ワード、あるいはグループをウインドウ内にリストするオプションが用意されており、学生が特定の問題に遭遇したとき、それと関連する音を練習できるようになっている。一例として、練習を必要とするエリアとして、例えば、“a l”をハイライトすると、この組合せを含むワードのリスト、例えば、

‘b a l k’, ‘t a l k’, および ‘w a l k’ が表示される。先生が模範を読み、学生は、これら単語を練習する。その後、学生は、練習中の元のサブワード、ワード、あるいはグループに戻り、練習を続ける。

【0067】学生は、教材を任意のモードにて復習することができる。これらモードには、例えば、先生の模範；模範とプロンプト；模範、プロンプト、録音；模範、プロンプト、録音、比較などのモードが含まれる。

【0068】学生の教材は、グラフィックイラストレーションにて提示することもできる。学生は詳細のためにズームインすることができる。学生は、教材を巡ってナビケートし、学習すべきエリアあるいは細部をハイライトし、選択することができる。学生のパフォーマンスを、テレスコープグラフィック表現にて提示し、学生が、終了したセッションの全てあるいは一部分にアクセスできるようにすることもできる。学生は、ズームインして、技能をさらに磨くことも、教材全体を復習することも、あるいは一部を選択することもできる。レベルが高くなるほどイラストは少なくなる。ズームインすると、より細かな部分がより詳細に示される。学生は、こ

の全体的なグラフィック表現から、どの教材から開始するかを決めることができる。

【0069】学生のパフォーマンスのスコアリングおよび評価に関しては、様々な技法および動作をシステムに組み込むことができる。スコアリング技法は当分野において周知である。ただし、一つの好ましい形態においては、自動化音声認識技術において周知のように、カスタム化されたスコアリングテーブルがキャリブレーションポイントとしての信頼スコアに基づいて生成される。例えば、ある言語における音素と二重母音の基本的項目を表す一連の文が提示され、学生がこれら文を読むと、学生の音声録音され、サブワードを解析することで、その学生のベースラインスコアあるいは開始パフォーマンスが決定される。これは、進歩を知るため、練習のレベルを決めるため、あるいは、練習すべきエリアを知るために用いられる。さらに、語学クラスの様々な等級レベルに対する基準スコアのテーブルが用意し、ある与えられた学生の集団から構成される語学クラスのグレードを評価することもできる。さらに、学生の進歩を、個人的に測定および評価し、他の学生と比較することもできる。

【0070】学生の音声のスコアは、サブワード単位であることも、ワード単位で得ることも、文あるいはパラグラフ単位で得ることもできる。さらに、学生は、総合スコアを得ることも、あるいは音声の各部分のスコアを得ることもできる。

【0071】音声の検証に関する正規化問題はインタフェースを通じて管理される。画面上のアニメ化されたカーソルを用いることで、システムによって、あるいは学生によって、様々な異なる継続期間と複雑さの会話を設定することができる。学生がアニメ化されたカーソルに沿って読み進むと、検証過程はハイライトされたテキストを解析されるべきサウンドファイルと比較する。

【0072】ある与えられた言語の既知のサブワードあるいはワードに対して、ある録音された音を自然な音を強調するためにさしはさむこともできる。これらワードは、事前に録音された辞書から取ることも、アプリケーションから取ることも、あるいは他のソースから取ることもできる。

【0073】得られたベースラインスコアは、テーブル（例えば、図1のテーブル52）内に記録される。テーブル52は、様々な形態を取ることができ、選択された学生に対するレッスンの適正なレベルや、評価方法を決めるために用いられる。このテーブルを利用することで、システムは、信頼尺度の閾値を変更することで、同一のテキストや内容を能力の異なる学生に対して用いることが可能となる。

【0074】学生は、さらに、グラフィカルユーザインタフェースを用いて、信頼尺度、等級レベル、その他の属性に対する閾値を設定あるいは修正することができ

る。自身の進歩を追跡するためには、学生によって識別番号が入力され、ベースラインスコアが評価され、その後のレッスンのスコアが評価され、スコアカスタム化されたレッスンが作成され、学生の進歩が追跡される。

【0075】図9は、学生がこの登録を行なうための一つの好ましい方法を示す。この方法900は、ステップ902において、学生が識別番号を入力することによって開始される。ステップ904において、学生の等級レベルが評価され、ステップ906において、スコアが記録され、ステップ908において、その後のレッスンが選択され、ステップ910において選択されたレッスンが評価され、ステップ912において、学生の記録が更新される。

【0076】図10は、ステップ904における学生の等級レベルを評価する過程を方法1000としてより詳細に示す。この過程のステップ1002において、第一のパラグラフが表示され、ステップ1004において、学生が第一のパラグラフを読み、ステップ1006において、システムによって信頼スコアが測定され、ステップ1008において、パラグラフ全体／パラグラフの要素に対する等級あるいはスコアが得られる。ステップ1010において、テーブルを検索することで、得られたスコアが他のスコアと比較され、ステップ1012において、所定の閾値を超えるか調べられる。閾値を超える場合は、第二のパラグラフが評価のために表示され、超えない場合は、ステップ1014において、単に等級レベルが表示される。

【0077】第二のパラグラフが表示された場合は、ステップ1016において、学生が第二のパラグラフを読み、ステップ1018において、信頼レベルが測定され、ステップ1020において、パラグラフ全体／サブパラグラフに対するスコアが得られる。ステップ1022において、再び、検索テーブルを用いて、等級あるいはスコアが決定される。これらステップが、ステップ1024において、検索テーブルから得られたスコアが所定の閾値を超えなくなるまで反復される。

【0078】図11は、ステップ908におけるレッスンの選択と910におけるレッスンのスコアリングに対するスコアリングの実例1100を示す。この一例としてのレッスンは、初級レベルの学生に対するもので、90%なるスコアを持つ。最初に、ステップ1102において、学生がある文を朗読するように求められ、この文の各単語には、図面に示すようなスコアが与えられる。各単語に対するこれらスコアから、学生に対して、どの単語に対するレッスンを与えるべきかが識別される。示される例では、単語“fox”は、たった50%なるスコアを受け、このため、ステップ1104において、単語“fox”に対してさらなるテストが行なわれる。次に、学生の単語“fox”の各文字の発音に対するスコアが求められる。説明の例では、ステップ1106にお

いて、“f”と“x”の音がさらなるレッスンを要求されるものとして決定され、ステップ1108において、レッスンのために、それぞれ、“f”と“x”の音を含む基本的な語彙集ワードが選択される。次に、ステップ1110において、ステップ1102における最初の文で低いスコアを受けた他のワードあるいは音（例えば、“jumps”、“lazy”、および“dog”）に対して、ステップ1104～1108と同一の動作が反復される。次に、ステップ1112において、各識別された音を含む様々な単語がレッスンとして訓練される。必要であれば、ステップ1114において、学生に対して、辞書からの録音された音がモデルサ音として再生される。次に、ステップ1116において、レッスンがスコアされ、レッスンの評価に対するテーブルが作成される。

【0079】本発明によるシステムおよび方法は、多くの好ましい機能およびアプリケーションを提供する。上述の様々な機能を組み合わせることで、学習のための機能に富むアプリケーションを生み出すことができる。このシステムには、システムの使用を容易にするために、台本に基づくレッスンの組合せが用意されている。ただし、経験を積むことで、先生あるいは学生は、特定の学生の学習問題あるいは学習形態に合わせて（およびPL94-142に準拠する個人的なプランを作成するために）機能の顧客に合わせて自由に組み合わせることができる。システムは、さらに学生のスコアとその学生の技能を向上させるために利用可能な機能に基づいて様々な機能の組合せを推奨する。システムは、学生が目標言語の発音の基礎を学習できるように工夫されたレッスンのファイルを備える。加えて、言語、文法、スペリング、シンタックス、および発音規則に対する参照テーブルが、それらと関連する補強学習のためのヘルプメッセージと共に用意されている。このシステムは、柔軟性に富み、学校教育にも、成人教育にも利用できる理想的なツールを提供する。

【0080】指導下での学習体験（Directed learning experience）—基本的なシステム機能は、テキストから音声に変換し、この音声を学生に再生することで、言語がどのような発音されるかを実際にやってみせることにある。こうすることで、学生は、学習すべき言語の音に慣れることができる。学生は模範を聴き、発音を学ぶことができる。

【0081】任意の単語を音声にて聴く（Listen to any word）—テキスト／音声変換技術を用いることで、学生は、ネイティブスピーカがいなくても、言語の音を真似ることができる。録音されたサンプルや教材などの入手やネイティブスピーカの個人的な確保には、例えば、英語を学習する学生にとって、多くの場合、かなりの制約があるが、テキスト／音声変換技術を用いることで、これら制約を除去することができ

る。ウェブ上の全ての材料、任意のテキストファイル、および準備された任意の教材をマルチメディア言語教材とすることができる。このシステムを用いれば、任意の自動的に生成されたテキストファイルを用いて、最新の言語教材を作り出すことができる。例えば、映画、テレビ、あるいはニュース番組から字幕テキスト(closed captioning text)を収集することで、システムの機能によって教材として用いることができるテキストファイルを作成することができる。

【0082】リスニングコンプリヘンション(Listening comprehension)ーテキストから音声への変換処理という基本的なシステム機能は、学生に、特別な教材を作成する必要も、他人に聴いてもらう必要もなく、リスニングコンプリヘンション技能を訓練する機会を提供する。この場合は、テキストは機能のパフォーマンスを向上させるために隠される。

【0083】模範とプロンプト(Example and prompt)ーテキストのテキスト/音声変換処理と顔のアニメーションとを組み合わせることで、学生に対して模範を示すことができる。システムのもう一つの機能によって、学生は、模範を復唱するように催促される。学生は、システムのこの機能を用いることで、模範を聴き、訓練することができる。この場合は、録音、等級付け、システムからのフィードバックなどは行なわれない。

【0084】模範、プロンプト、録音(Example, prompt, record)ーシステムは、これら3つの機能を組合せることで、学生が模範を聴き、いつなにか読むかのプロンプトを聴きあるいは見、サブワード、ワード、フレーズなどを発音する自身の努力(音声)を録音するための手段を提供することができる。

【0085】模範、プロンプト、録音、再生(Example, prompt, record, playback)ーシステムは、これら機能を組み合わせることで、学生が、模範を聴き、プロンプトを聴くか見るか、音声を録音し、模範と学生の音声を学生によって比較できるようにサイド・バイ・サイドに再生するための手段を提供することができる。

【0086】自分で選択できる補強(Self-selected reinforcement)ー学生が特定の録音されたサンプルに関して問題を見け、ヘルプが必要であると判断した場合、学生は、機能セクションのワークスペースセクション内に表示されるコンテキストに特定なヘルプにアクセスすることができる。学生は、ハイライトされた単語と関連する言語の規則を識別するヘルプシステムにアクセスすることができる。このシステムは、“Try this... (これを試して下さい)” シリーズを、学生一般あるいはその学生が属する集団にしばしば認められる既知のエラーに基づいて所定の順番にて提出する。学生は、これら補強ヘルプメッセ

ージの幾つかあるいは全てを、見たり、聴いたりすることができる。

【0087】模範、プロンプト、録音、再生、比較、結果の表示(Example, prompt, record, playback, compare, displays results)ーシステムの諸機能を総合的に活用する例として、音声と視覚を伴う先生の模範、先生の表情を変化させるオプション、先生の音声特徴を変化させるオプション、プロンプトの聴覚あるいは視覚を通じての提供、音声の録音、パフォーマンスを聴けるようにするための再生、自動音声認識技術を用いての学生の単語の発音の処理、パフォーマンスに関する数値尺度の計算、フレキシブルで適応可能な基準の枠組み内でのパフォーマンスの表示などの機能を組合せることもできる。

【0088】文法スキル(Grammar skills)ーワード処理要素(プログラム)の追加により、語学の先生は、文法スキルを教えたり、補強したりすることができる。学生は、テキスト文と質問を聴き、答えを作成し、答えをシステムに話すか文字にて入力する。すると、ワード処理プログラムの結果として、文の構文エラーなどのサンプルが作成される。システムは、これを用いて言語の既知の規則に基づいて推奨される“Try this...”のサンプルを作成する。問題領域が上述のようにハイライトされ、学生はワーキングウィンドウを用いて技能を訓練する。システムには、アジアの言語を母国語とする学生が英語を学習する際の典型的な発音問題に関する教材が用意されている。

【0089】システム内の複数の機能を組み合わせることで学生に学習教材を提供することができる。複数の機能を組み合わせることで、先生は、学習の計画、教育の方法、および学生へのフィードバックを管理することができる。このように、このシステムは、柔軟性が極めて高く、ユーザはこのシステムを自在に活用することができる。先生にとっては、PL94-142に準拠した個々の学生にあった個人的な学習計画を容易に作成することができる。このシステムの重要な機能は、テキストから音声への変換と顔のアニメーションを組み合わせ、発音を視覚的に支援できることである。こうして、教材からの実例を用いて発音と関連する典型的な、顔、口、舌の動きを解説することができる。この機能は、母国語以外の言語を学習する学生にとって重宝であるばかりか、音声セラピストを目指す学生にとっても有益である。

【0090】語学の学習する際には、専門的な関心あるいは主題の内容が必要とされる場合がある。例えば、自動車の部品を扱う会社の従業員や、医療機関の従業員は、会社の資料からの教材をリスニングの練習に用いたい場合がある。このような場合、ユーザは、本発明のシステムを用いることで、ユーザが置かれている環境においてよく耳にし、従って、理解したいと考える専門用語や語句を練習することができる。

25

【0091】上述の説明は、単に本発明の特定な実施例を開示するためのものであり、本発明をこれらに限定することを目的とするものではない。従って、本発明は上述の実施例に制限されるものではなく、当業者においては本発明の範囲から逸脱することなく様々な代替の実現が可能であると思われる。

【図面の簡単な説明】

【図1】本発明によるシステムの略図である。

【図2】本発明による対話型学習のために生成されるウィンドウを示す図である。

【図3】本発明による方法の概要を示す流れ図である。

【図4】本発明による方法におけるテキストを選択し、可聴音声を作成する過程の詳細な流れ図である。

【図5】本発明による方法におけるテキストを選択し、アニメーションと可聴音声を作成する過程の詳細な流れ図である。

【図6】本発明による方法における録音する過程の詳細な流れ図である。

【図7】本発明による方法における録音のもう一つの方法の詳細な流れ図である。

【図8】本発明による方法における再生過程の詳細な流れ図である。

【図9】本発明による方法における学生が登録する過程を示す流れ図である。

【図10】本発明による方法における等級レベルの評価（音声部分の評価）の過程を示す流れ図である。

【図11】本発明の方法によるスコアリングの実例を示す流れ図である。

【符号の説明】

10 対話型語学指導システム

12 コンピュータシステム

26

14 マイクロプロセッサ

16 入力デバイス

18 マイクロホン

20 入力リンク

22 ディスプレイデバイス

24 音声スピーカ

26 出力ファイルインタフェースユニット（メモリ）

30 テキスト／音声変換（TTS）モジュール

32 自動音声認識（ASR）モジュール

32a 音声情報検証（VIV）部分

36 規則モジュール

38 辞書あるいは録音された音声ファイル

40 レッスンファイル

34 アニメーションモジュール

42 ワークスペースモジュール

44 発音モジュール

46 マッピングテーブル

48 ワードプロセッシングモジュール

50 録音／再生モジュール

20 60 ウィンドウ

62 ワークスペース

64 フィードバックエリア

66 アニメーションエリア

68 制御エリア

68a 音量調節

68b 速度調節

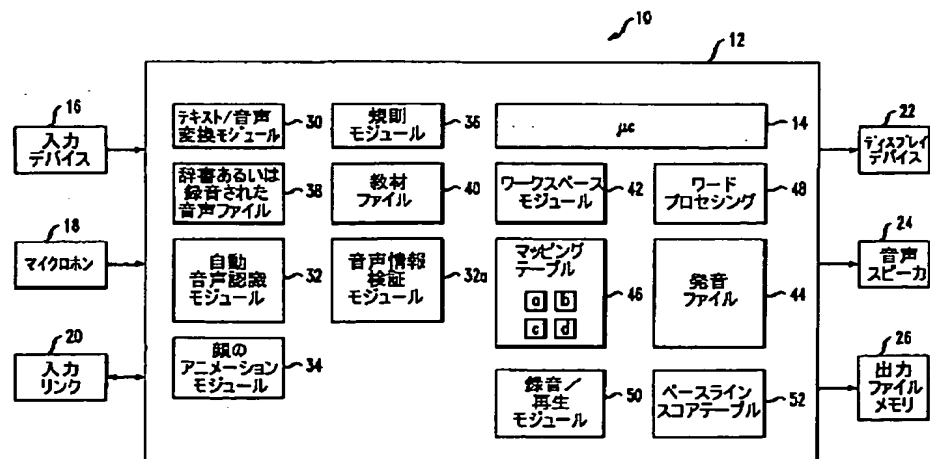
68c 停止ボタン

68d 再生ボタン

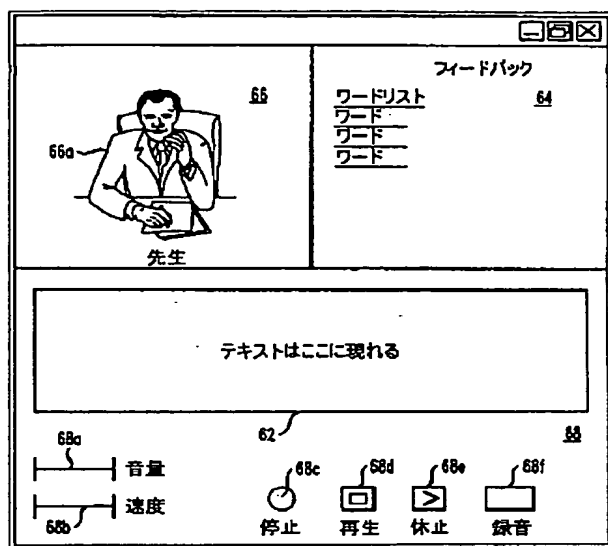
68e 休止ボタン

30 68f 録音ボタン

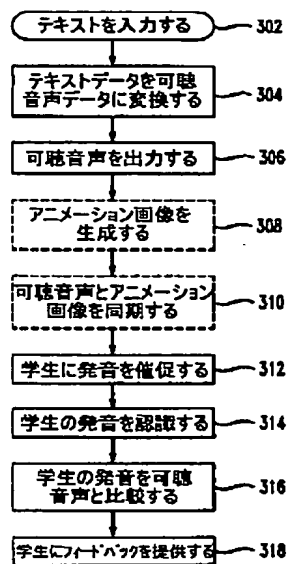
【図1】



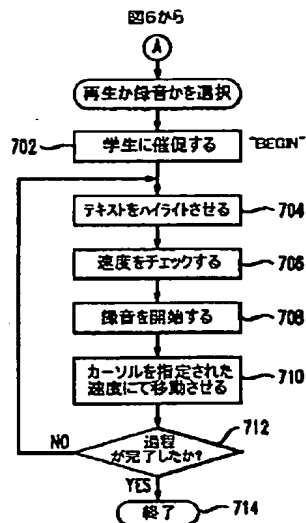
【図 2】



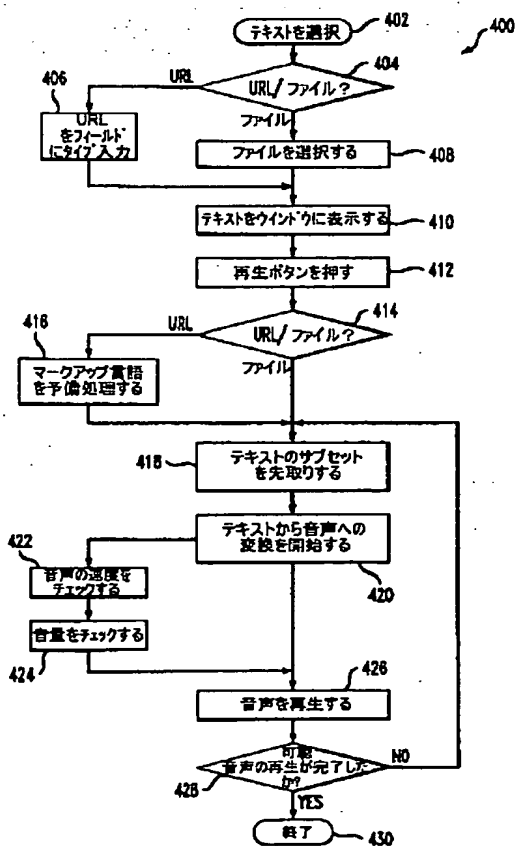
【図 3】



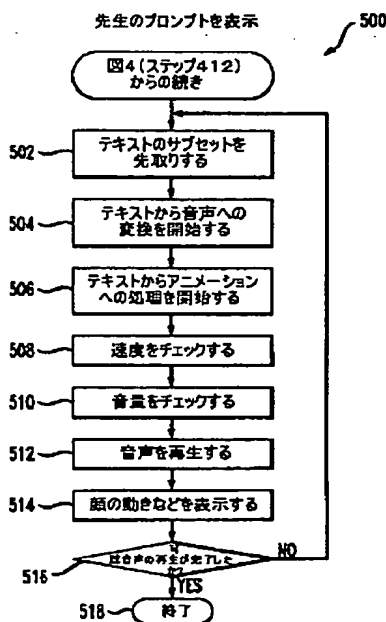
【図 7】



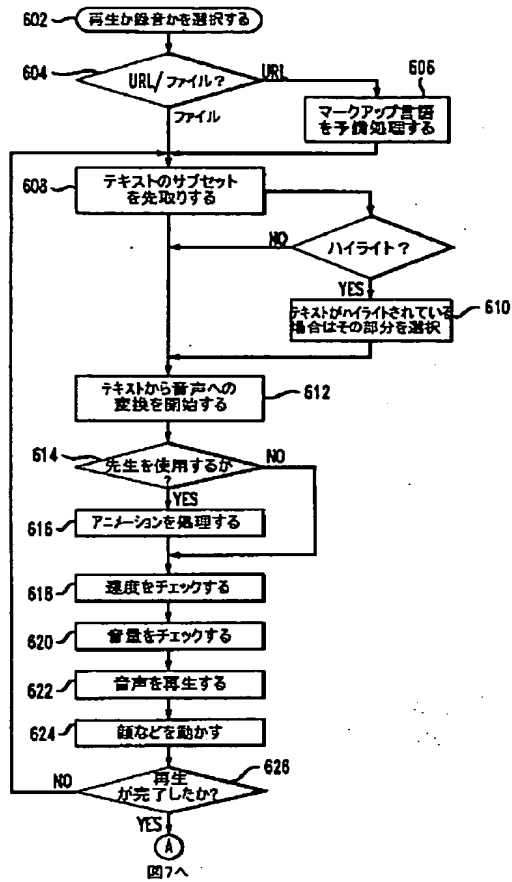
【図 4】



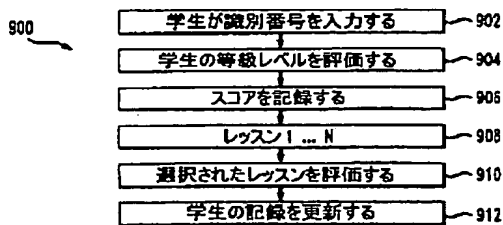
【図 5】



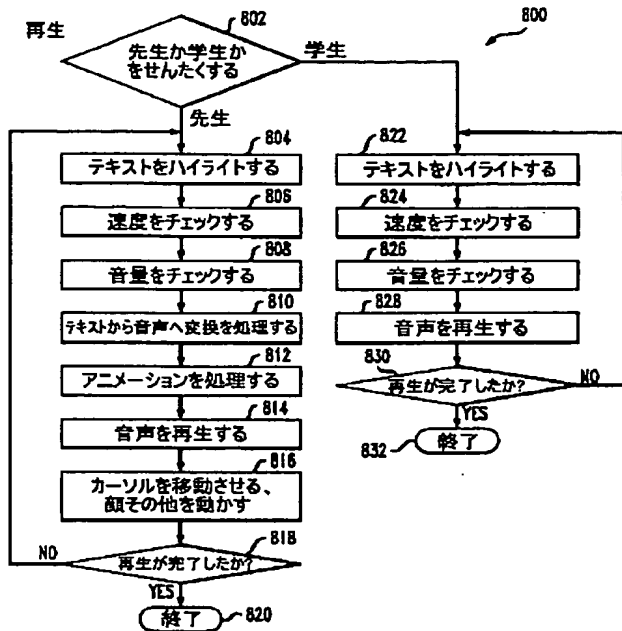
【図 6】



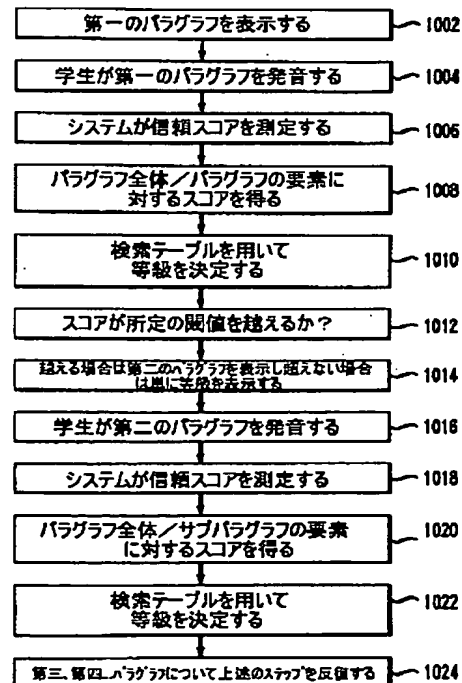
【図 9】



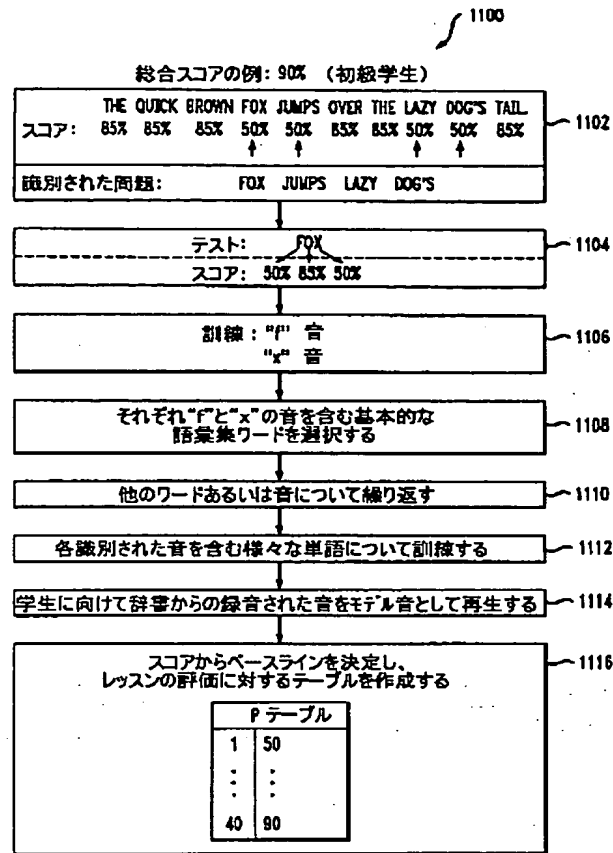
【図 8】



【図 10】



【図11】



フロントページの続き

(51) Int. Cl. 7

識別記号

F I

ターマート* (参考)

G 0 9 B 5/14
15/00
19/04
19/06
G 1 0 L 13/00
21/06
15/00
15/28

G 0 9 B 5/14
15/00
19/04
19/06
G 1 0 L 3/00

R
S
5 5 1 E
5 7 1 H

(72) 発明者 ナディン ブラックウッド
アメリカ合衆国 07747 ニュージャージー
イ, マタワン, アバーデーン ロード 33
(72) 発明者 キー ビー, リ
アメリカ合衆国 07974 ニュージャージー
イ, ニュープロヴィデンス, ラニーメド
パークウェイ 225

(72) 発明者 マイケル マックナーニー
アメリカ合衆国 07728 ニュージャージー
イ, フリーホルド, セッター プレイス
159

(72) 発明者 チーリーン シー
アメリカ合衆国 07922 ニュージャージー
イ, パークレイ ハイツ, マックマン ア
ヴェニュー 150

(72) 発明者 アラン チャンドラセカラン サレンドラ
ン
アメリカ合衆国 08904 ニュージャージー
イ, ハイランド パーク, セダー レーン
88シー

(72) 発明者 ジアリン ソング
アメリカ合衆国 07922 ニュージャージー
イ, パークレイ ハイツ, サウス, ホリー
グレン レーン 49

(72) 発明者 キル ゴー
アメリカ合衆国 07076 ニュージャージー
イ, スコッチ ブレインズ, パイン スト
リート 558